Contents

Statistics

G.N.Dar

May 12, 2020

1 Chi Squre Distribution

Before we start with chi-squre disrtibution directly, we are going to understand the background of it. This distibution is used to check the theoritical hypothesis against the experimental result,and conclusions are drawn weather the experiment is unbiased really random or the biased one so that acceptance or rejection of the experimental observation is done.

In this distribtion the important parameter is degree of freedom (*df*) say for example in a fair dice throw it is 5, in coin toss it is 1 soon.By *df*, we mean independency, for example throwing a dice 100 times we ,5 outcomes say $1, 2, 3, 4, 5$ are independent but 6th will be dependent. That is if $1, 2, 3, 4, 5$ appears $14, 16, 15, 19, 14$ times then 6 th will be restricted to 20times only. Similarly in tossing a coin 100 times if head appears 55 times then tail has to appear only 45 times and soon. We define degree of freedom as $df = number of catagories - 1$ say in a die it is $df = 6 - 1 = 5$ and in coin it is $df = 2 - 1 = 1$.

Randome variable in Chi Squre Distribution

If $X_1, X_2, X_3, X_4, X_5, ...X_n$ are the random variables and each following a normal distibution or standard normal distribtions then the variable χ^2 defined as below is following a Chi-squrre distribtion.

$$\chi^2 = Y = X_1^2 + X_2^2 + X_3^2 + X_4^2 + ... + X_n^2$$

support base of Y

$$R_y \in [0, \infty]$$

$$\chi^2(y) = \begin{cases} Cy^{\frac{n}{2}-1)}e^{\frac{-1}{2}y}, & \text{if } \chi^2 \in R_x \\ 0, & \text{otherwise} \end{cases}$$

c is a contant caontaining Gamma funciton.

for example in a 500 dice throw each independent variable say $1, 2, 3, 4, 5$ following a normal distribtion then the sum of squre of each of these variables follows a Chi-squre distibution.

i)

$$\chi^2 \geq 0$$

ii) Chi-squre is a skewed distibution for lower degree of freedom and peaks immedielty near zero but for larger degree of freedom > 90, it is approaching to normal distribtion with mean at df and varaince $\sqrt{2df}$

1.1 Applications of Chi Squre Distribution

It has a large number of Applications and some are as
i) Chi-squre test of goodness of fit
ii) Chi-squre test for independence of attributes
iii) Chi-squre test for the population varaince.

Chi-squre test of goodness of fit It is a tool or test to describe the magnitude of discrepancy between theory and observation. It enables us to find if the deviations of observations from theory is just by chance or due to inadequacy of the theory to fit the obseved data. This test is knowns as χ squre test of goodness of fit.

If $O_i(i = 1, 2, 3...n)$ is observed data and $E_i(i = 1, 2, 3, 4..n)$ are the corresponding expected data (theoritical frequncy) then the statistic χ^2 defined as

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

with

$$\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i$$

following a chi squre distribtion with $(n - 1)$ degrees of freedom.

Acceptance or rejection of the hypothesis is done based on the comparision of this statistic with table value (given by R.A.Fisher avalible for various levels of confidence ordinariliy upto 30 degree of freedom). if it is less than the table value , hypothesis is accepted else rejected.

Example: A die thrown 132 times and results are

Number turned up:123456 with frequncy as frequncy : 162025142928 respectively. test the hypothesis that the die is unbiased.

solution based on hypothesis the expected: $\frac{132}{6} = 22$

O (Observed)	E (Expected)	(O-E)^2	$\frac{(O-E)^2}{E}$
16	22	36	1.64
20	22	4	0.18
25	22	9	0.41
14	22	64	2.91
29	22	49	2.23
28	22	36	1.64
			$\sum \frac{(O-E)^2}{E} = 9.01$

here $df = 6 - 1 = 5$

for 5 degrees of freedom at 5% level of significance, the table value is $\chi^2 = 11.07$ the calculated value is less than the table and hence there is no evidence against the hypothesis.

Example 2

Check for a die thrown 300 times with obseved values as $1 = 48, 2 = 52, 3 = 55, 4 = 45, 5 = 53, 6 = 47$ times

1.2 Derivation of Chi Squre distribtion for one degree of freedom

Let z is a standard normal variable with mean 0 and varaince 1 that is $z\ S(0, 1)$. Then

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Let X is a random variable of χ^2 distribtion we have by definition

$$X = z_1^2 = z^2$$

$$z = \sqrt{x}$$

for x we know about the distribtion funciton

$$x < 0, F_x(x) = 0$$

therefore a probability function can of X can be derived by diffrentiating the distribtion funciton as

$$f_x(x) = \chi_x^2(x) = \frac{d}{dx} F_x(x)$$

$$f_x(x) = \chi_x^2(x) = \frac{d}{dx} \int_{-z}^{z} f_z(z)dz) = \frac{d}{dx} \int_{-\sqrt{x}^{\sqrt{x}} f_z(z)dz)}$$

$$f_x(x) = \chi_x^2(x) = f_z(z)\frac{d}{dx}(+z) - f_z(-z)\frac{d}{dx}(-z)$$

$$f_x(x) = \chi_x^2(x) = f_z(\sqrt{(}x)\frac{d}{dx}(\sqrt{(}x) - f_z(-\sqrt{(}x)\frac{d}{dx}(-\sqrt{x})$$

Since $f_z(z)$ is given to be normal distribtion funciton(putting it value) and after diffrentiating part gives us $\frac{1}{\sqrt{x}} = x^{-\frac{1}{2}}$ we arrive at

$$\chi_x^2(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{1}{2}x}$$

OR

$$\chi_x^2(x) = \frac{1}{\sqrt{2\pi}} x^{\frac{1}{2}-1} e^{-\frac{1}{2}x}$$

knowing the Gamma funciton$\Gamma(\frac{1}{2}) = \sqrt{\pi}$ we get

$$\chi_x^2(x) = \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} x^{(\frac{1}{2}-1)} e^{-(\frac{1}{2}x)}$$

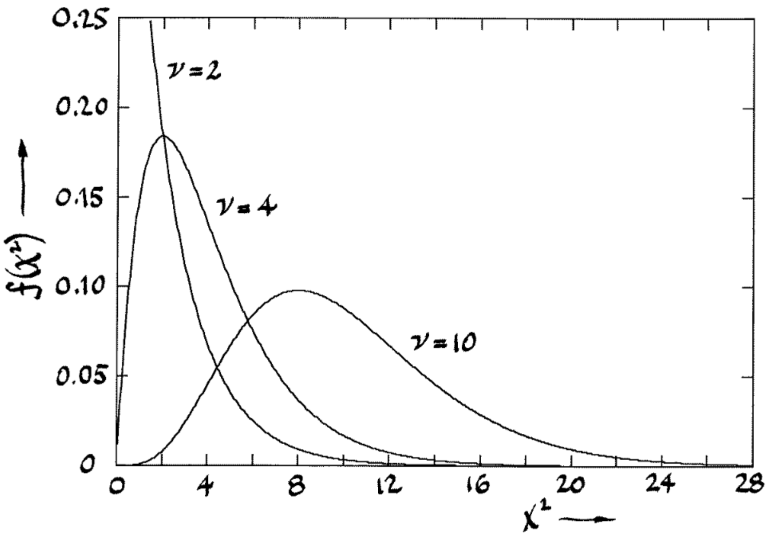

Figure 1 — The chi-square distribution for $\nu = 2$, 4, and 10.

which is the required chi squre distibution for one degree of freedom. For n degree of freedom we have

$$\chi_x^2(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{(\frac{n}{2}-1)} e^{-(\frac{1}{2}x)}$$

n is known as shape parameter and 2 scale parameter

varying value of n define the shape of distibution. Dependening on degree of freedom we have right skewed and normalized chi squre distibutions.

Alternative Method

If X follows standar normal distibution with mean 0 and varaince 1

$$XSN(0,1)$$

then the pdf for X^2 will be derived as

$$f(x) = \frac{d}{dx} P(X^2 < x)$$
$$= \frac{d}{dx} P(-\sqrt{x} < X < \sqrt{x})$$
$$= \frac{d}{dx} \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

for dummy index we have

$$= \frac{d}{dx} \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2}} d\mu$$
$$= \frac{1}{\sqrt{2\pi}} \frac{d}{dx} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{\mu^2}{2}} d\mu$$

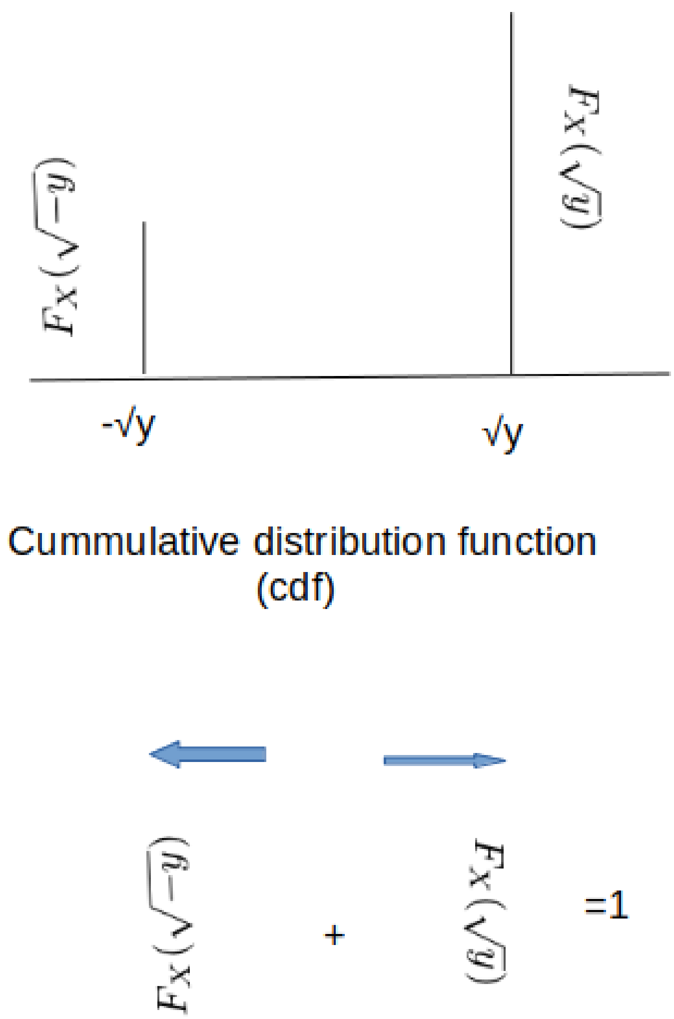

Figure 1: graphical understanding

$$= \frac{2}{\sqrt{2\pi}} \frac{d}{dx} \int_0^{\sqrt{x}} e^{-\frac{\mu^2}{2}} d\mu$$

using standard integral method we have

$$f(x) == \frac{e^{-\frac{x}{2}} x^{-\frac{1}{2}}}{\sqrt{2}\Gamma(\frac{1}{2})}$$

in standard notaiton we write

$$\chi_x^2(x) = \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} x^{(\frac{1}{2}-1)} e^{-(\frac{1}{2}x)}$$

1.3 Expected value of Chi Squre Distribution

If x follows a χ^2 distribtion then

$$E(X) = \int_0^\infty x\chi^2(x)dx$$

$$E(X) = \int_0^\infty xCx^{(\frac{n}{2}-1)} e^{(-\frac{1}{2}x)} dx$$

$$E(X) = C\int_0^\infty x^{(\frac{n}{2})} e^{(-\frac{1}{2}x)} dx$$

$$E(X) = C[-x^{\frac{n}{2}} 2e^{(-\frac{1}{2}x)}]_0^\infty + C\int_0^\infty \frac{n}{2} x^{(\frac{n}{2}-1)} 2e^{(-\frac{1}{2}x)} dx$$

do some calculations and see we get back a distribtion which gives probability from all over the support base as 1

$$= n\int_0^\infty \chi(x)dx = n.1 = n$$

hence expected value is the degree of freedom

1.4 Variance of Chi Squre Distribution

we know

$$\nu_2 = V(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2$$

therefore

$$E(X^2) = \int_0^\infty x^2 Cx^{(\frac{n}{2}-1)} e^{(-\frac{1}{2}x)} dx$$

$$E(X^2) = C\int_0^\infty x^{(\frac{n}{2}+1)} e^{(-\frac{1}{2}x)} dx$$

$$E(X^2) = C[-x^{(\frac{n}{2}+1)} e^{(-\frac{1}{2}x)}(2)|_0^\infty + 2\int_0^\infty ((\frac{n}{2}+1)) x^{\frac{n}{2}} e^{(-\frac{1}{2}x)} dx]$$

$$E(X^2) = C(n+2)\int_0^\infty x^{\frac{n}{2}} e^{(-\frac{1}{2}x)} dx$$

$$E(X^2) = C(n+2)\left[-2x^{\frac{n}{2}} e^{(-\frac{1}{2}x)} \mid_0^\infty + \int_0^\infty 2\frac{n}{2} x^{(\frac{n}{2}-1)} e^{(-\frac{1}{2}x)} dx\right]$$

$$E(X^2) = C(n+2)\left[0 - 0 + n\int_0^\infty x^{(\frac{n}{2}-1)} e^{(-\frac{1}{2}x)} dx\right]$$

$$E(X^2) = Cn(n+2)\int_0^\infty x^{(\frac{n}{2}-1)} e^{(-\frac{1}{2}x)} dx$$

$$E(X^2) = Cn(n+2)\int_0^\infty \chi(x) dx$$

keeping in view the total probability equal to 1 we have

$$E(X^2) = n(n+2).1 = n^2 + 2n$$

therefore

$$V(x) = \sigma^2 = n^2 + 2n - n^2 = 2n$$

OR

$$\sigma = \sqrt{2n}$$

therefore we conclude to say that Chi squre distibution has one important parameter that is degree of freedom and the mean and varaince is degree of freedom dependent.

Home Work: Derive an expression for Chi squre distribtion of degree 1

NOTE : You need to gain more understanding on this topic by going beyond this elementary introduction